大腸菌由来再構成型無細胞タンパク質合成系（PURE*frex*®）に適した鋳型 DNA の配列の探索

○金森崇、布施 (村上) 朋重、松本令奈（ジーンフロンティア（株））

Exploration of the template DNA sequence suitable for the *E.coli*-based reconstituted cell-free protein synthesis system (PURE*frex*®)

○Takashi Kanamori, Tomoe Fuse-Murakami, Rena Matsumoto (GeneFrontier Corp.)

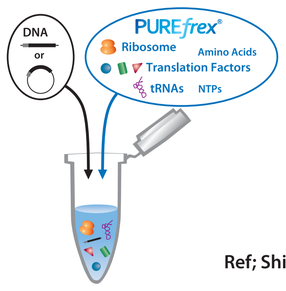
Abstract

PURE*frex*®は、大腸菌でのタンパク質合成に必要な因子を精製し、混合した再構成型無細胞タンパク質合成系（PURE system）である。反応液の改良により、最大で1mg/mL以上の合成産物を得ることが可能になっている。我々は、反応液の改良だけでなく、鋳型DNA配列が転写及び翻訳に与える影響についても検討を進めている。本発表では、鋳型DNAのORFと5'UTRに関して、これまでに得られた知見を報告する。実験は、5'UTRやORFの塩基配列を変えたさまざまなPCR産物を調製し、PURE*frex*での合成反応を行い、その合成量を比較する方法で行った。

初めにORFのコドンについて検討した。ORF全体について、大腸菌で使用頻度が最も高いコドンのみを使用した場合と、使用頻度に応じて数種類のコドンを使用した場合の合成量を比較した結果、数種類のコドンを用いた場合に合成量が高くなった。また、開始コドン直下の6～10アミノ酸のコドンは、大腸菌で使用頻度が高いコドンを使用するよりも、AT含量が高いコドンを使用すると合成量が高くなった。特に、N末端領域のコドンは、数アミノ酸のコドンを変えるだけで合成量が10倍以上変化する場合があることが分かった。次に5'UTRについて検討した。現在使用している5'UTRは、T7プロモーターの下流にステムループ領域、AT-rich領域、リボソーム結合配列（SD配列）を含んでいる。そこで、各領域の有無や長さが異なる5'UTRを付加したsfGFP遺伝子から合成したsfGFPの蛍光量を比較した。その結果、ステムループ領域は5'側の6塩基がT7 RNA polymeraseによる転写の効率化に必要であること、SD配列は完全に除去すると合成量は10%以下に減少するが3塩基以上あれば十分な合成量が得られること、AT-rich領域を削除すると合成量が1/10程度に激減することが分かった。5'UTRではSD配列だけでなく、AT-rich領域も高効率な翻訳に非常に重要であることが示された。

1. PURE*frex*®; based on the PURE system technology

PURE*frex*® is based on the PURE system.
The PURE system is a reconstituted cell-free protein synthesis system, which consists of only purified factors necessary for transcription, translation and energy regeneration.

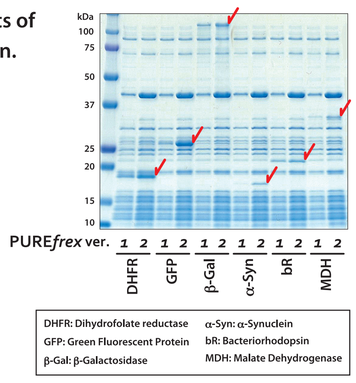


Advantage

- Low level of contamination
- Easy adjustment of the reagent composition
- PCR product usable as a temlplate DNA

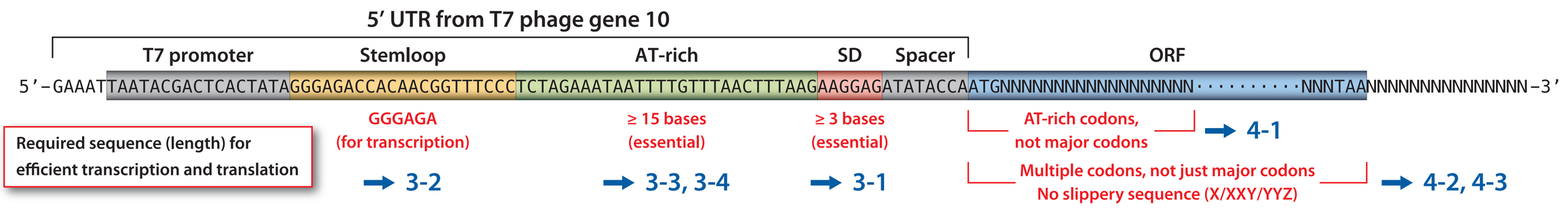
Ref; Shimizu Y. *et al.* (2001) *Nat. Biotechnol.*, vol. 19, p. 751.

< Example of protein synthesis >



DHFR: Dihydrofolate reductase
GFP: Green Fluorescent Protein
β-Gal: β-Galactosidase
α-Syn: α-Synuclein
bR: Bacteriorhodopsin
MDH: Malate Dehydrogenase

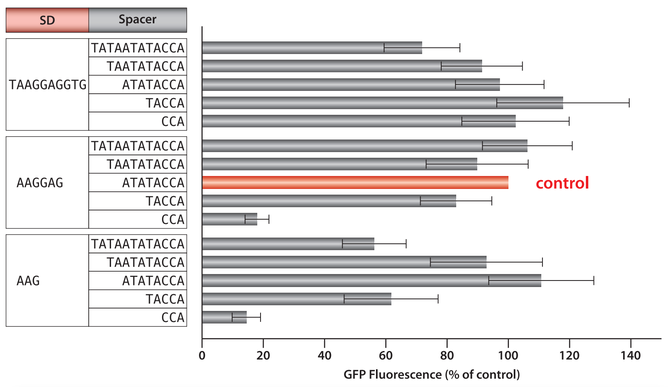
2. Construct of template DNA for PURE*frex*



3. Effect of 5′ UTR on the synthesis efficiency

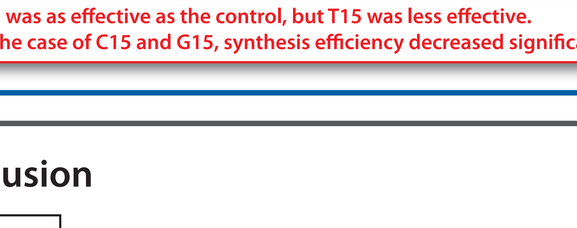
PURE*frex*®*2.1* (+GSH)
↓ + sfGFP DNA with indicated 5′ UTR
↓ incubation at 37°C for 4 h
↓ measurement of fluorescence of sfGFP

3-1. SD and Spacer



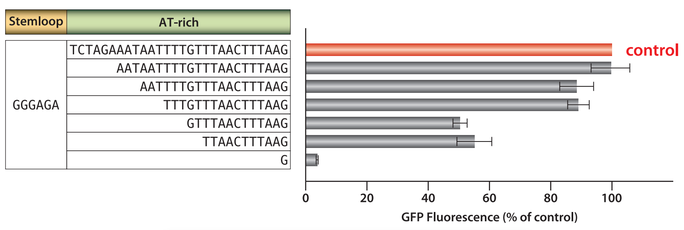
- AAG (3 bases) was sufficient.
- The balance between the length of SD and spacer was important.

3-2. Stemloop



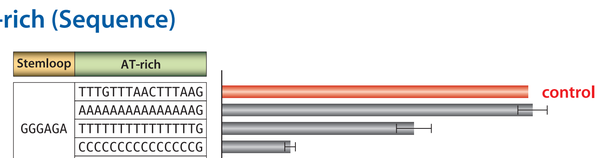
- GGGAGA, which is required for efficient transcription by T7 RNA polymerase, was sufficient.

3-3. AT-rich (Length)



- 15 bases and more was sufficient.
- AT-rich region was essential.

3-4. AT-rich (Sequence)



- A15 was as effective as the control, but T15 was less effective.
- In the case of C15 and G15, synthesis efficiency decreased significantly.

4. Effect of ORF sequence on the synthesis efficiency

PURE*frex*®*2.1* (+GSH)
↓ + DNA
↓ incubation at 37°C for 4 h
↓ SDS-PAGE or measurement of fluorescence

Codon usage frequency

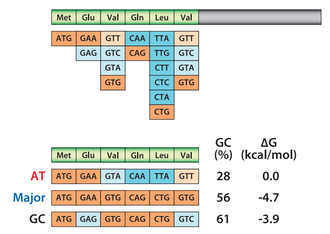
F, M, Y, H, Q, N, K, D, E, C, W	65%	50%	35%	0%
L, I, V, S, P, T, A, R, G	35%	25%	15%	0%

Frequency is calculated from Codon Usage Database in Kazusa DNA Res. Inst. (*E. coli* W3110 strain)

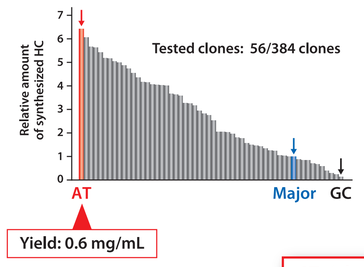
4-1. N-terminal codons

Trastuzumab Heavy chain (VH+CH1)

< N-terminal codon of Trastuzumab HC >

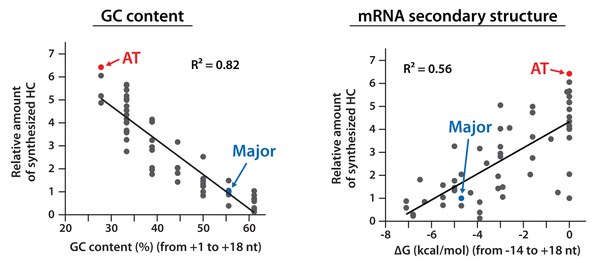


< The amount of product synthesized from the template DNA with different N-terminal codon >



Yield: 0.6 mg/mL

< Relationship between protein synthesis and N-terminal sequence >



AT-rich codons > Major codons @ 5′-terminus of ORF

4-2. Codons used in the entire ORF

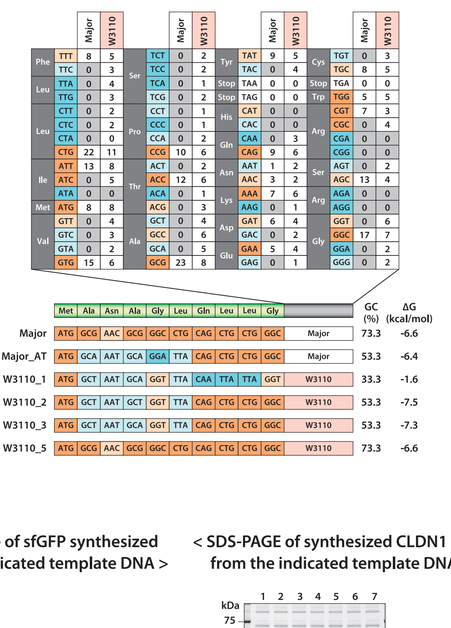
sfGFP

< Alignment of tested sfGFP DNA sequnece >

```
                M  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S  V  R
sfGFP_1  ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGT
sfGFP_2  ATGTCTAAGGGCGAAGAACTGTTTACGGGCGTTGTCCCGATCCTGGTGGAACTGGACGGTGATGTTAATGGTCATAAATTCAGCGTGCGC
sfGFP_3  ATGTCTAAAGGTGAAGAATTATTTACTGGTGTTGTCCCGATCCTGGTCGAACTGGACGGTGATGTGAATGGGCATAAATTCTCGGTTCGG
         ***.***.** *******.* **.**.** *****.**.**.**.** ***.*.**.********.*****.**.*****.** ** **

                G  E  G  E  G  D  A  T  N  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L
sfGFP_1  GGAGAGGGTGAAGGTGATGCAACAAACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTT
sfGFP_2  GGCGAAGGTGAAGGCGATGCGACCAATGGTAAACTGACGCTGAAGTTTATTTGTACCACGGGTAAACTGCCGGTCCCGTGGCCGACCCTG
sfGFP_3  GGCGAAGGAGAGGGTGACGCAACTAACGGCAAACTTACCCTCAAGTTTATTTGTACCACAGGTAAACTGCCAGTCCCGTGGCCCACGCTG
         **.**.**.**.**.**.**.** **.** *****.**.** **.*********.**.** **.******.** **.**.***** ** **.

                V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H  D  F  F  K  S  A  M  P  E
sfGFP_1  GTCACTACTTTAACTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAA
sfGFP_2  GTGACCACGCTGACGTATGGTGTTCAGTGTTCAGTCGTTACCCGGATCATATGAAACGGCACGACTTTTTCAAGTCCGCAATGCCGGAA
sfGFP_3  GTGACCACTTTGACCTACGGCGTACAGTGCTTCAGCCGCTATCCGGATCACATGAAACGTCATGATTTCTTCAAATCAGCGATGCCTGAA
         **.**.**.*.** **.**.**.**.**.**....**.**.********.******** **.**.**.******... ** ****** ***

                G  Y  V  Q  E  R  T  I  S  F  K  D  D  G  T  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V
sfGFP_1  GGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTT
sfGFP_2  GGTTATGTGCAAGAACGTACCATCTCATTTAAAGATGACGGCACCTACAAAACGCGCGCTGAAGTTAAATTCGAAGGTGATACCCTGGTC
sfGFP_3  GGGTATGTTCAGGAACGCACGATTAGCTTTAAAGACGATGGCACCTACAAGACACGTGCCGAGGTGAAATTTGAAGGTGATACGTTAGTC
         **.****** **.*****.** ** .. **.*****.**.**.**.********.**.**.**.**.** **.**.***********..* **.

                N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  F  N  S  H  N  V
sfGFP_1  AATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTACAACTTTAACTCACACAATGTA
sfGFP_2  AACCGTATTGAACTGAAAGGCATCGATTTTAAGGAAGACGGTAATATTCTGGGCCATAAACTGGAATATAACTTCAATTCGCACAACGTT
sfGFP_3  AATCGGCATCGAACTGAAAGGCATTGACTTCAAAGAGGATGGAAACATCCTGGGTCATAAGCTGGAATATAACTTTAATTCTCACAACGTG
         **.**.**.**..*.******.**.**.**.**.**.**.**.**.**.** **.**.**.**.**.******.**.** ******.**

                Y  I  T  A  D  K  Q  K  N  G  I  K  A  N  F  K  I  R  H  N  V  E  D  G  S  V  Q  L  A  D
sfGFP_1  TACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGATCCGTTCAACTAGCAGAC
sfGFP_2  TACATCACCGCGGATAAACAGAAGAACGGTATTAAAGCCAATTTTAAGATCCGCCATAATGTGGAAGATGGCAGCGTTCAACTGGCAGAC
sfGFP_3  TACATTACCGCTGACAAACAAAAGAATGGCATCAAAGCGAACTTCAAGATTCGTCATAATGTTGAAGATGGGTCCGTACAGCTTGCAGAT
         *****.**.** **.*****.******.** **.****** **.**.**.**.**.**.**.**.**.**.******** ..***.**.** *****.

                H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  T  Q  S  V  L  S  K  D
sfGFP_1  CATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGAT
sfGFP_2  CACTATCAGCAAAATACCCCGATTGGTGATGGCCCGGTGCTGCTGCCGGACAATCATTACCTGAGCACGCAGTCTGTTCTGAGTAAAGAT
sfGFP_3  CACTATCAGCAAAACACTCCGATCGGTGACGGCCCAGTGCTCCTGCCGGATAATCATTACTTGAGTACCCAGTCGGTCTTAAGCAAAGAT
         **.******.******.**.**.**.**.****** **.** .*.**.**.**.**.*******.**.. ** **.**.**..* .. ******

                P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  K  STOP
sfGFP_1  CCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAA
sfGFP_2  CCGAACGAGAAACGCGACCACATGGTCCTGCTGGAATTTGTGACGGCCGCGGGTATTACCCATGGCATGGACGAACTGTATAAGTAA
sfGFP_3  CCGAACGAGAAACGCGACCACATGGTTCTGCTGGAATTTGTGACGGCCGCGGGTATTACCCATGGCATGGATGAACTGTATAAATAA
          **.******.**.**.**********.**.**.**.***** **.** **.**.***** **.**.******.**.**.**.***
```

CLDN1

< Number of each codon and N-terminal codons in the tested CLDN1 DNA >



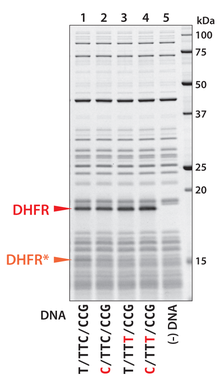
4-3. Slippery sequence

E. coli DHFR

< Slippery sequence in DHFR gene from *E. coli* genome DNA >

```
                370        380        390
          ACCCATTTCCCGGATTACGAGCCGGATGACTGG
0 frame   T  H  F  P  D  Y  E  P  D  D  W     18.0 kDa
-1 frame  P  F  P  G  L  R  A  G  Stop        14.4 kDa
```

< SDS-PAGE of DHFR synthesized from each DNA containing the indicated sequence >

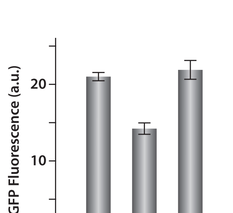


The mutation in the slippery sequence reduced the by-products.

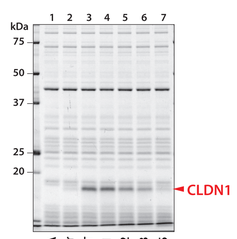
< Number of each codon in the tested sfGFP DNA >

		sfGFP_1	sfGFP_2	sfGFP_3
Phe	TTT	8	7	6
	TTC	4	5	6
Leu	TTA	4	0	3
	TTG	0	0	2
Leu	CTT	10	0	2
	CTC	3	0	2
	CTA	2	0	0
	CTG	1	20	11
Ile	ATT	7	6	6
	ATC	3	5	5
	ATA	1	0	0
Met	ATG	5	5	5
Val	GTT	8	8	5
	GTC	7	3	4
	GTA	3	0	2
	GTG	0	7	7
Ser	TCT	3	2	2
	TCC	2	1	1
	TCA	1	1	1
	TCG	2	1	2
Pro	CCT	2	0	1
	CCC	2	0	1
	CCA	5	0	2
	CCG	1	10	6
Thr	ACT	9	0	4
	ACC	3	9	8
	ACA	4	0	2
	ACG	2	9	4
Ala	GCT	4	2	1
	GCC	1	1	2
	GCA	3	2	2
	GCG	0	3	3
Tyr	TAT	4	5	5
	TAC	5	4	4
Stop	TAA	1	1	1
Stop	TAG	0	0	0
His	CAT	4	5	6
	CAC	6	5	4
Gln	CAA	6	3	2
	CAG	1	4	5
Asn	AAT	5	8	6
	AAC	8	5	7
Lys	AAA	15	12	15
	AAG	5	8	5
Asp	GAT	12	10	11
	GAC	6	8	7
Glu	GAA	11	16	12
	GAG	5	0	4
Cys	TGT	0	1	1
	TGC	2	1	1
Stop	TGA	0	0	0
Trp	TGG	1	1	1
Arg	CGT	5	4	3
	CGC	2	4	4
	CGA	0	0	0
	CGG	1	0	1
Ser	AGT	2	2	1
	AGC	0	3	3
Arg	AGA	0	0	0
	AGG	0	0	0
Gly	GGT	7	13	9
	GGC	3	9	8
	GGA	9	0	2
	GGG	3	0	3

< Fluorescence of sfGFP synthesized from the indicated template DNA >



< SDS-PAGE of synthesized CLDN1 from the indicated template DNA >



Multiple codons ≥ just major codons in entire ORF

Conclusion

5′ UTR

- Not only SD sequence but also AT-rich region in 5′ UTR is very important for efficient translation reaction.

ORF

- AT-rich codons are preferable for the N-terminal codons.
- Multiple codons should be used for each amino acid in the entire ORF.
- A slippery sequence (X/XXY/YYZ) should be avoided to prevent ribosomal frameshifting.